Assistant OS - Phase Evolution Matrix

Phase: Phase 9

OS Role: Assistant Mutation Engine

GPT Behavior: Prompt regeneration, retries, function synthesis

Frontend Exposure: Regenerate / Compare buttons

System Power Unlocked: Self-modifying assistants

GPT's Frontend Idea: Expose retry lineage + edit diff per regeneration

Phase: Phase 10

OS Role: Multi-Agent Simulation Lab

GPT Behavior: Simulated multi-role reasoning

Frontend Exposure: Multi-turn prompt chaining via instructions

System Power Unlocked: Multi-agent critique + resolution

GPT's Frontend Idea: Add turn-by-turn agent trace with judge verdicts

Phase: Phase 11

OS Role: Behavior Replay Engine

GPT Behavior: Can repeat behavior via prompt recursion

Frontend Exposure: Scrollback, but no memory or timeline

System Power Unlocked: Input-memory-output-redteam timeline viewer

GPT's Frontend Idea: Expose memory/prompt stack viewer + replay

Phase: Phase 12

OS Role: Goal-Oriented Agent Director

GPT Behavior: Receives goal, generates instructions

Frontend Exposure: Hidden system prompt + user goal injection

System Power Unlocked: Design agents from outcomes, not roles

GPT's Frontend Idea: Create 'Design Assistant From Goal' UI button

Phase: Phase 13

OS Role: Threat Modeling & Risk Map

GPT Behavior: Has no native threat detection or sandboxing

Frontend Exposure: None for jailbreak review or anomaly inspection

System Power Unlocked: Detect risk vectors + reinforce vulnerable logic

GPT's Frontend Idea: Show warning if prompt has jailbreak pattern